12th IFAC International Symposium on
Advanced Control of Chemical Processes
July 14-17, 2024. Toronto, Canada

Systematic Selection of Constraints for a Novel Dynamic Flux Balance Model of Mammalian Cell Cultures

A. Ghodba, * **A. Richelle,** ** **P. Agarwal,** *** **C. McCready,** ***
L. Ricardez-Sandoval, * **H. Budman,** *

* *Chemical Engineering Department, University of Waterloo, Waterloo, Canada (Email: hbudman@uwaterloo.ca)*
** *Sartorius Corporate Research, Brussels, Belgium*
*** *Sartorius Corporate Research, Toronto, Canada*

Abstract: The dynamic flux balance model (DFBA) is a constrained-based optimization modeling approach that has gained popularity for describing microbial cultures but has not been thoroughly investigated for mammalian cell cultures due to their relative complexity. This research aims to identify a DFBA model with minimal constraints and associated parameters to predict data for the fed-batch operation of a mammalian CHO cell culture. The Bayesian Information Criterion (BIC) is used to find a minimal set of kinetic constraints. The resulting DFBA model is used to predict 24 metabolites, biomass, and titer with 85 parameters that has a lower BIC and higher R^2 as compared to previously reported kinetic models.

Keywords: Dynamic Flux Balance Analysis; CHO cell; Bayesian Information Criterion

1. INTRODUCTION

Monoclonal antibodies (mAb) are recombinant proteins used to treat different diseases ranging from autoimmune diseases to different types of cancer. The global market for antibodies is expected to surpass 300 billion US dollars by 2025, which is up to 50% of the market's size of biotherapeutics in 2020 (Lu et al., 2020). Chinese Hamster Ovary (CHO) cells are the most popular microorganism for the generation of monoclonal antibodies. CHO cells are utilized extensively because they can adapt to various growth mediums, produce larger titers than other mammalian cell lines, and are biosafe host cells that are less vulnerable to viral infection (Fox et al. (2004)). The abrupt increase in demand for mAbs, partly due to an increasingly aging population, has motivated the industry to further improve the manufacturing processes of mAbs. Mathematical models are crucial for achieving this goal since they provide a quantitative representation of the underlying processes, allowing for a systematic and rigorous analysis of the system without the need for costly and time-consuming experimentation. This can lead to significant cost savings in terms of materials, time, and resources. Different types of models can be considered for describing bio-processes including CHO cell cultures. Dynamic flux balance model (DFBA) is increasingly becoming the method of choice for modeling cell cultures because it constrains the metabolic network of a given microorganism based on an a priori known metabolic network of reactions. Hence, it may require fewer parameters compared to other less structured modeling approaches that do not impose such constraints, e.g., kinetic and data-driven models.

The DFBA model assumes that the cell acts as an optimizing agent that distributes resources among different biochemical pathways to accomplish a particular biological objective, e.g., maximization of growth and ATP, or minimization of metabolic burden. This optimization is subject to constraints such as stoichiometric relations, positivity concentration of metabolites, and reaction rate constraints (Carvalho et al. (2019)). As the DFBA model involves the solution of constrained optimization problems, the choice and calibration of the objective function and the metabolic constraints are crucial for fitting the data while avoiding the over-fitting of noisy data. In principle, for DFBA (Eq. 1) there is no need to calibrate the kinetic parameters for each possible reaction involved in the metabolic network, but only for a subset of them. Thus, only parameters related to metabolites involved in limiting constraints need to be estimated whereas the time evolution of non-limiting metabolites is expected to follow stoichiometric relations among metabolites. Hence, the key challenge is the identification of the minimum set of limiting constraints that best describe the data, thus avoiding over-parametrization and over-fitting.

While DFBA has been successfully applied to describe dynamic microbial populations, e.g. Ecoli, B. Pertussis (Budman et al. (2013)), its application to CHO cells in a fed-batch operation has not been studied in literature due to the relatively high complexity of biochemical phenomena involving mammalian cultures, e.g. programmed cell death and a significantly more complex media composition. This work presents a DFBA model for a fed-batch culture of CHO cells with a small number of calibration parameters to fit the data while avoiding over-fitting. Finding a small number of constraints and associated tuning parameters is crucial to capitalize on the potential ability of DFBA to generate more compact models as compared to other modeling approaches. Note that an insufficient number of

parameters can lead to a lack of fitting. Thus, a trade-off between the number of parameters and model accuracy must be made. This balance is evaluated in this study using the Bayesian Information Criterion (BIC) (Burnham and Anderson (2004)). This criterion combines a goodness-of-fit term and a penalty term to prevent over-fitting. Hence, it provides a balance between having enough parameters to accurately describe the data (sensitivity) while avoiding over-fitting of data or inferring relationships that do not exist (specificity). Avoiding over-fitting is crucial if the model is to be used for optimization that will potentially require predictions for operating conditions that were not considered during model training.

Although the DFBA model has been recently applied to CHO cells by our research group (Nikdel et al. (2018)), that study required the solution of a challenging optimization problem that involved the gradual addition of constraints until fitting was satisfactory, which resulted in a large number of model parameters. In contrast, in this work, a minimum set of constraints is identified from an initial large number of assumed constraints based on the minimization of the BIC. The resulting model results in good training and prediction accuracy with a better BIC as compared to the initial model with all possible constraints.

Following the above, the goals of this paper are two-fold: i- a methodology to reduce the number of constraints of the model and associated parameters so as to avoid over-fitting, and ii- the development of a novel DFBA model for mammalian cultures based on the identified constraints.

2. MATHEMATICAL FRAMEWORK

In this section, the approach for selecting a minimum set of constraints is presented. The approach involves the following 3 elements: i- development of a DFBA model, ii- DFBA parameter estimation based on data using a bi-level optimization, and iii- Selection of a small set of constraints based on the Bayesian Information Criterion. Each one of these elements is discussed below.

2.1 Dynamic Flux Balance Analysis

The DFBA problem is formulated as a sequence of discrete linear optimization (LP) problems as described in Eq.(1).

$$\begin{aligned}
\underset{v_k}{max} \quad & \mathbf{c}^T \mathbf{v}_k && \text{(1a)}\\
s.t. \quad & \mathbf{S}.\mathbf{v} = 0 && \text{(1b)}\\
& (1-e)\mathbf{f}(\psi_{k-1}, \Theta) \leq \mathbf{v}_{exchange} \leq \\
& (1+e)\mathbf{f}(\psi_{k-1}, \Theta) && \text{(1c)}\\
& \mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max} && \text{(1d)}\\
& \psi_k = \psi_{k-1} + X_{v,k-1}\mathbf{v}_{exchange,k-1}\Delta t + F_k \Delta t \\
& + \Delta V \psi_{k-1} \geq 0 && \text{(1e)}
\end{aligned}$$

where k indicates the number of sampling time interval; $\mathbf{S}$ is the stoichiometric matrix; n_r is the number of reactions in the metabolic network; $\mathbf{c} \in \mathbb{R}^{n_r}$ is a biological objective coefficient vector; $\mathbf{v}_k \in \mathbb{R}^{n_r}$ is the flux vector at time interval k and subscript *exchange* refers to exchange reactions in the metabolic network; $\psi_k \in \mathbb{R}^{n_m \times n_k}$ is metabolite concentration vector at the time interval k; $X_{v,k}$ is viable cell density at time interval k. The concentrations' dependent functions $\mathbf{f}(.)$ represent kinetic constraints, capacity constraints, and other biological constraints; e, to be referred to as the coefficient of variation, is used to account for the uncertainty of measurements. This coefficient can be adjusted such that the LP admits a feasible solution at each measurement time t_k, i.e. to ensure that the constraints can be fulfilled at each time instant. Since fed-batch operation is being considered, $F \in \mathbb{R}^{n_m \times n_k}$ describes a bolus addition of each metabolite at time k. Also, the volume of the reactor decreases over time because samples are regularly taken out from the reactor for measurement. Thus, ΔV is introduced to account for these volume changes. Δt denotes the length of each time interval.

The optimization problem (1) includes different types of constraints including i- mass balances of intra-cellular metabolites, which are assumed to follow a quasi-steady state response (1b), ii- concentration-independent lower and upper bound of each flux (1d), and iii- positivity of concentration of metabolites (1e) and iv- concentration dependent kinetic constraints that are imposed on exchange fluxes to further regulate the dynamic behavior (1c). Constraints (1b) and (1d) cannot be tuned since the former solely depends on stoichiometry while the latter is determined by the maximal and minimal specific consumption/production rates observed in the data for the entire set of experimental conditions. Hence, the only parameters that are available for tuning are those involved in concentration-dependent constraints (1c) which are generally expressed by standard kinetic forms such as Monod or Hill functions. The kinetic constraints assumed in the current study are presented in Appendix A.

2.2 Parameter Estimation

A parameter estimation procedure is performed to identify the kinetic parameter values in Eq.(1c), which results in the best fit of model predictions to the experimental data. This task is accomplished by minimizing the normalized sum of squared errors (SSE) between the model prediction and experiment measurements (Eq.(2)) with respect to parameters' values subject to the satisfaction of the LP problem in Eq.(1). Normalizations of the terms included in the SSE are used since metabolite measurements are not within the same range of values. This parameter estimation problem is formulated as a bi-level optimization (Eq.(2)) whose outer layer calculates the SSE whereas the inner layer satisfies the DFBA problem in Eq.(1). The decision variables are the tuning parameters for the outer problem and the fluxes for the inner layer where the latter is solved with the parameters proposed by the outer layer for each iteration during the optimization search.

$$\begin{aligned}
\Theta = \quad & arg \underset{\Theta}{min} \ \frac{1}{n}\sum_{k=1}^{n_k} w_k \left\| \mathbf{y}_p(t_k) - \mathbf{y}(\Theta, t_k) \right\|^2 && \text{(2a)}\\
& s.t. \quad Problem\ (1) && \text{(2b)}
\end{aligned}$$

As shown in Eq.(2), $\Theta \in \mathbb{R}^{n_\Theta}$ are parameters in the kinetic constraints of the model (1c), n is the number of measurements, and n_k represents the number of time intervals. Also, $\mathbf{y}_p \in \mathbb{R}^{n_k \times n_y}$ denotes the plant measurements, whereas $\mathbf{y} \in \mathbb{R}^{n_k \times n_y}$ represents the DFBA model

predictions. $w_k \in \mathbb{R}^{n_y}$ are the weights used to normalize the errors in different metabolites.

2.3 Reduction of the model parameters based on the Bayesian Information Criterion

Once the parameters in the metabolic constraints are identified, the minimum set of metabolic constraints for predicting the cell culture behavior must be determined. One approach to determining minimum constraints is finding active constraints in the model by solving Lagrangian multipliers. However, because of the multiplicity in the model, this approach is not effective. Also, the Lagrange multipliers' approach does not explicitly consider the number of parameters used in constraints that can serve as an indicator of potential over-fitting. Therefore, in this work, the minimum number of parameters is determined based on the minimization of BIC which balances the accuracy of prediction for all metabolites with the number of parameters. BIC was selected since it is recommended for small data sets and larger models. If the prediction accuracy of a particular metabolite is poor, e.g. due to multiplicity of the LP solution, it will introduce a kinetic constraint for that metabolite. Also, by explicitly considering the number of parameters the BIC penalizes over-parameterization. The BIC is defined as follows:

$$BIC = ln\,(n)\,p + nln\left(\widehat{\sigma_e^2}\right) \tag{3}$$

where

$$\widehat{\sigma_e^2} = \frac{1}{n}\sum_{k=1}^{n_k}\left(y_p(t_k) - y(t_k)\right)^2 \tag{4}$$

where p is the number of parameters used in the model.

To find the minimal number of constraints among all the constraints considered in the earlier parameter estimation problem (2), an additional bi-level optimization is formulated. In this bi-level optimization, the outer level involves the minimization of the BIC with respect to different sets of kinetic constraints (1c), whereas the inner layer is the LP defining the DFBA. The outer layer of the proposed bi-level optimization, shown in Eq.(5), results in a MIP problem with respect to different sets of constraints. The BIC minimization is performed to achieve a trade-off between the number of constraints and associated parameters and the fit to data.

$$\min_{x_i} \quad ln\,(n)\,p + nln\left(\widehat{\sigma_e^2}\right) \tag{5a}$$
$$s.t. \quad for\ t = 1, ..., n_k \tag{5b}$$
$$\max_{v_k} \quad \mathbf{c}^T\mathbf{v}_k \tag{5c}$$
$$s.t. \quad Eq.(1b - 1e)$$
$$p = \sum_{i\in I}\mathbf{w}_i\mathbf{x}_i \tag{5d}$$
$$\mathbf{x}_i \in \{0,1\}\ \forall i \in I \tag{5e}$$

where I denotes a set of constraints, each one penalized by a weight $\mathbf{w}_i$ that is directly proportional to the total number of parameters involved in the corresponding set of kinetic constraints, and a set of decision binary variables $\mathbf{x}_i$ that is equal to 1 if ith constraint is selected, or 0 if it is not.

2.4 Metabolic network and stoichiometric matrix

The accuracy and reliability of DFBA predictions depend on the quality and completeness of the metabolic network used for the analysis as well as the constraints imposed on it. The network must be validated prior to model training to ensure that the reactions and pathways are biologically relevant and consistent with experimental data. The metabolic network used for this study is the genome-scale metabolic network of CHO Cell metabolism performed by Hefzi et al. (2016) which consists of 4456 metabolites and 6663 reactions.

In order to assess the ability of this metabolic network to predict the experimental observations, the feasible solution flux space is explored with the Flux Variability Analysis (FVA) method, which is a method used to determine the range of fluxes' values that satisfy, within some tolerance, the original FBA problem. Also, FVA can be used to find a suitable variation coefficient for Eq.(1c). The FVA method can be formulated as a series of LP problems, i.e.,

$$\max_{\mathbf{v}}/\min_{\mathbf{v}} \quad \mathbf{v}_i \tag{6a}$$
$$s.t. \quad \mathbf{S}.\mathbf{v} = 0 \tag{6b}$$
$$\mathbf{c}^T\mathbf{v} \geq \mu Z_0 \tag{6c}$$
$$(1-e)\mathbf{v}_{exp,exchange} \leq \mathbf{v}_{exchange} \leq (1+e)\mathbf{v}_{exp,exchange} \tag{6d}$$
$$\mathbf{v}_{min} \leq \mathbf{v} \leq \mathbf{v}_{max} \tag{6e}$$

where Z_0 is the biological objective function that is determined from the experimental measurements. $\mathbf{v}_{exchange}$ are exchange fluxes and $\mathbf{v}_{exp,exchange}$ are experimental fluxes of exchange reactions. As shown in Eq.(6), the FVA method involves the solution of $2n_r$ LPs subject to a constraint that either considers suboptimal solutions ($\mu < 1$) or it enforces the exact optimality of the FBA problem ($\mu = 1$). To examine the metabolite network, Algorithm 1 is used in combination with experimental data. This algorithm calculates the feasible solution range of each exchange flux which can then be compared with the fluxes calculated from data. Calculations are started with $\mu = 1$, and $e = 0.001$ as the best possible scenario. However, if some or all of the experimental fluxes' values are not within the calculated bounds, these initial parameters are changed so that the feasible solution will include the experimental values. Then, based on the values of μ and e that are required to bound all the measured fluxes, the ability of the metabolic network to explain the data is evaluated.

2.5 Solver

The solver "cplexlp" (Dual-simplex algorithm), from IBM ILOG CPLEX was used to solve the linear problem of the DFBA model, and the KNITRO solver from Artelys was used to solve the non-linear problems. The processor of the system is 12^{th} Gen Intel(R) Core(TM) i7-12700 and is equipped with 16GB of RAM.

3. CELL CULTURE PROCESS

The experimental data used to train and validate the DFBA model were measured for 24 metabolites, biomass,

Algorithm 1 Proposed approach to evaluate the feasibility of the CHO metabolic network

Require: Experimental measurements for all metabolites and Z_0
1: initialize with $\mu = 1$, and $e = 0.001$
2: calculate fluxes at all time intervals for all metabolites based on the experimental measurements
3: **for** $i = 1, ..., n_m$ **do**
4: set sufficiently large bounds for the exchange reaction of metabolite i in Eq.(6d)
5: impose experimental fluxes $\pm$ e% for exchange reactions of other metabolites in Eq.(6d)
6: find the minimum and maximum feasible bounds for exchange flux i using Eq.(6)
7: **if** experimental exchange flux i is in feasible bounds **then**
8: go to step 12
9: **else**
10: increase e or decrease μ and repeat from step 3
11: **end if**
12: compare the experimental flux with the feasible range of flux values for exchange flux i at all times
13: **end for**

and titer in four fed-batch CHO cell cultures. All cultures were performed by CHO-DG44 cell line, producing IgG, at 36.8 C, pH=7.1, in AMBR 250 bioreactor. The bolus was added to all cultures on days 3, 4, 5, 6, 7, 8, 9, 10, and 11. All experiments are carried out in the same way but glucose feeding varies between the cultures. In all batches, the same proprietary media from Sartorius was used. The initial cell density for all fed-batches was about $4 \times 1e5$ cell/mL and experiments were run for 12 days. The BioProfile FLEX analyzer was used to quantify total and viable cell density, glucose, glutamine, glutamate, ammonia, and lactate concentrations, whereas the NMR method was used to measure amino acid concentrations. Samples were taken three times per day and data were used to calibrate and validate the DFBA model.

4. RESULTS

As a first step, the feasibility of the metabolic network was evaluated by the FVA method (Algorithm 1). The analysis shows that for $e = 0.05$ and $\mu = 1$ a feasible solution resulted for all exchange fluxes, i.e. all the measured fluxes were within the bounds calculated from FVA analysis. These results indicated that the iCHOv1 metabolic network is an acceptable metabolic network to predict the experimental data of this study since only a 5 percent error is needed to bound the measured fluxes $e = 0.05$ which matches the expected value of error in these measurements.

The next step involves finding the kinetic parameters of metabolic constraints according to the procedure in (2). Two fed-batch cultures were used for calibration in this step. While for most DFBA models for bacteria, the DFBA is formulated in terms of maximization of growth rate, such an objective did not provide good results for the current study with CHO cells since the metabolic network allocates all resources to growth thus resulting in zero mAb productivity. A possible explanation for this result is that the cells under study are engineered by pool selection to produce a maximal amount of protein (mAb) but not necessarily for maximization of growth. Accordingly, the maximization of mAb production was chosen as the biological objective function for the DFBA model.

Table 1. Comparison of BIC in DFBA model for training and validation data sets

DFBA with	Training	Validation
	Fed-batch 1/2	Fed batch 3/4
25 kinetic constraints	-2312/-1794	-1924/-1778
15 kinetic constraints	-2397/-1901	-1964/-1874

The DFBA model was trained by using runs 1 and 2 and validated by runs 3 and 4. As illustrated in Fig.1, the model fits well with experimental data, and it effectively predicts the consumption and production of metabolites and mAb over time. Also, the growth and death phases of biomass have been accurately calculated. The SSE for training and validation data sets are 0.03 and 0.04, respectively. The resulting similar SSE for calibration and validation indicates the suitability of the model.

This initially calibrated model imposes 25 kinetic constraints on all metabolites of interest and biomass and has a large number of parameters (116). Our goal is to use models that require fewer parameters to avoid overfitting. To reduce the number of constraints and associated parameters, the bi-level optimization presented in section 2.3 was implemented, and the results are given in Table (1). The results in Table (1) indicate that the DFBA model with the selected subset of constraints has a lower BIC for both the training and validation data sets, which shows the reduced model is a superior model. As explained above, these BIC values are affected by the number of model parameters and the error in the prediction. The number of parameters in this reduced model decreased from 116 for the original model to 85. On the other hand, the BIC provides a trade-off between the number of parameters and the prediction for the metabolites related to the constraints that were pruned from the original model as shown in Fig.2. Although the error of prediction for the selected model with 85 parameters is 23% higher as compared to the model with 116 parameters, the BIC clearly favors the model with fewer parameters.

Due to the results of the optimization problem 5, the metabolites that are/are not involved in the dynamic constraints are as follows:

- **constricted metabolites:** Ala, Amm, Asn, Asp, Glc, Gln, Glu, Lac, Gly, Phe, Leu, Ser, Thr, Tyr, Biomass
- **non constricted metabolites:** Arg, For, His, Ile, Lys, Met, Pro, Succ, Trp, Val, mAb

Among the restricted metabolites, three metabolites, i.e. lactate, glucose, and biomass, were forced to be constricted. In fact, it is evident from the data that there is a metabolic shift from a state of high glycolysis flux and high lactate production in the earlier stages of the experimental runs to a state of low glycolysis flux and lactate consumption at later stages (see lactate plot in Fig. 2). The DFBA model could not capture these shifts unless kinetic constraints were added for glucose and lactate. Also, DFBA cannot inherently predict cell death since the

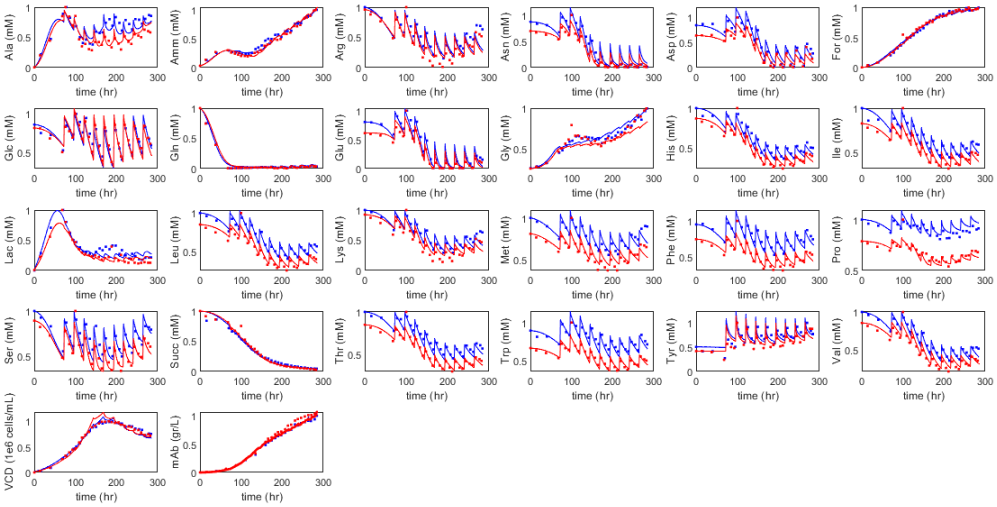

Fig. 1. Calibration and validation results of the DFBA model for fed-batch systems with normalized units

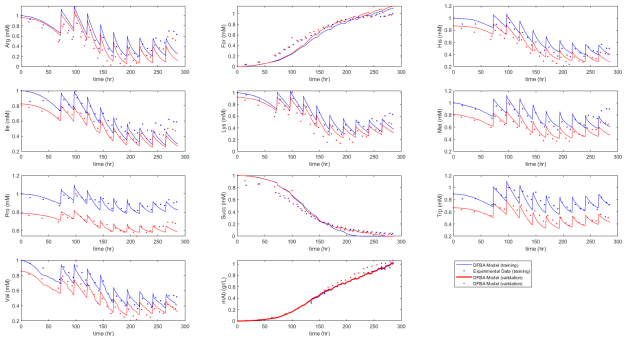

Fig. 2. Calibration and validation results of the DFBA model selection based on BIC for fed-batch systems with normalized units for non-constricted metabolites

mechanisms of death are not implemented in the metabolic network. To model death, a dynamic balance equation for biomass was developed (see biomass equation in Appendix A) that was based on growth-promoting compounds such as glucose and glutamine, toxic compounds such as lactate and ammonia, and cell death by autophagy that results from glucose deprivation.

Fig.3 illustrates the effect of removing/adding more constraints from/to the optimal point. Initially, two metabolites are randomly selected between constricted metabolites to investigate. When Leucine kinetic constraint is removed from the DFBA model, it results in a 20 unit decrease in the penalty term associated with the number of model parameters. However, it simultaneously leads to a notable 60 unit increase in the penalty term of the likelihood function. Similarly, the removal of Glycine is aimed at reducing the number of kinetic constraints to 13. Nonetheless, this results in a substantial increase in the penalty term for the likelihood function. This is because inaccuracies in Glycine predictions have the potential to affect the kinetic constraints associated with Ammonia and Serine due to high interaction among these metabolites. Conversely, the introduction of additional kinetic constraints from non-constricted metabolites, such as Methionine, Proline, and Tryptophan, does not significantly reduce the penalty term for the likelihood function. This is primarily due to their already robust predictive capabilities. However, the penalty term for the number of parameters increases.

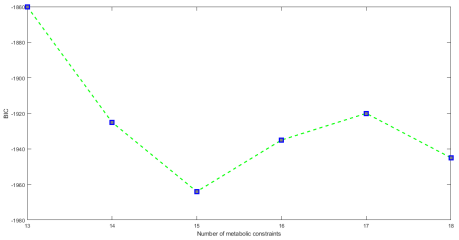

Fig. 3. Variations of the BIC around the optimal point

To assess the ability of the proposed model with 85 parameters to generate reliable predictions, this selected DFBA model is compared with a previously reported kinetic model for CHO cells Hille (2018). Hille's model was developed using a traditional approach of formulating a system of dynamic mass balances for each measured metabolite by assuming a priori knowledge of main reactions among metabolites and where each reaction is modeled by a Michaelis-Menten expression. This kinetic model involves 71 parameters and it is used to predict 23 metabolites, biomass density, and mAb. For the given data the BIC of Hille's kinetic model was equal to -170 which is significantly higher than the DFBA model in this work. Hence, this work demonstrates for the first time that a DFBA model that is developed based on a systematic selection of constraints provides a good trade-off between the number of parameters and accuracy as compared to models that were developed with traditional approaches that are not based on constrained optimizations. Furthermore, this result indicates that the explicit enforcement of detailed stoichiometry relations and positivity constraints of fluxes and concentrations are instrumental in avoiding model over-parameterization.

5. CONCLUSION

While DFBA models have been extensively studied for bacteria and algae, it has not been thoroughly applied to mammalian cell cultures due to the inherent complexity of mammalian cells as compared to other microorganisms.

To avoid overfitting of data it is crucial to identify which kinetic constraints must be included within DFBA to result in reasonable fitting of data while avoiding excessive over-parameterization. This work proposes a systematic approach to find these constraints based on the BIC measure that explicitly trades off fitting accuracy versus the number of model parameters. The identified constraints are used to develop a comprehensive model of a mammalian cell culture operated initially in batch and then in fed-batch operation. Since the BIC only assesses the model's quality in relation to other models, other metrics were also calculated to ensure the model's quality. For example, the residual error for the validation data set of the DFBA model is 86%, which indicates that the selected reduced model has good accuracy.

Furthermore, this work presents a first comparison between a DFBA of CHO cell culture to another model previously reported that was developed with traditional approaches of dynamic mass balances of key metabolites. The comparison of the BIC for the two models clearly showed the advantage of DFBA versus the traditional approach. Benders Decomposition method (Liñán and Ricardez-Sandoval, 2023) will be further investigated for solving (Eq.(5)).

ACKNOWLEDGEMENTS

We are grateful to Sartorius and Mitacs for their financial support.

REFERENCES

Budman, H., Patel, N., Tamer, M., and Al-Gherwi, W. (2013). A dynamic metabolic flux balance based model of fed-batch fermentation of bordetella pertussis. *Biotechnology progress*, 29(2), 520–531.

Burnham, K.P. and Anderson, D.R. (2004). Multimodel inference: understanding aic and bic in model selection. *Sociological methods & research*, 33(2), 261–304.

Carvalho, M., Nikdel, A., Riesberg, J., Lyons, D., and Budman, H. (2019). Identification of a dynamic metabolic flux model for a mammalian cell culture. *IFAC-PapersOnLine*, 52(1), 88–93.

Fox, S.R., Patel, U.A., Yap, M.G., and Wang, D.I. (2004). Maximizing interferon-γ production by chinese hamster ovary cells through temperature shift optimization: Experimental and modeling: Experimental and modeling. *Biotechnology and bioengineering*, 85(2), 177–184.

Hefzi, H., Ang, K.S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C.A., Baycin-Hizal, D., Huang, Y., Ley, D., et al. (2016). A consensus genome-scale reconstruction of chinese hamster ovary cell metabolism. *Cell systems*, 3(5), 434–443.

Hille, R. (2018). Run-to-run optimization of biochemical batch processes in the presence of model-plant mismatch.

Liñán, D.A. and Ricardez-Sandoval, L.A. (2023). A benders decomposition framework for the optimization of disjunctive superstructures with ordered discrete decisions. *AIChE Journal*, 69(5), e18008.

Lu, R.M., Hwang, Y.C., Liu, I.J., Lee, C.C., Tsai, H.Z., Li, H.J., and Wu, H.C. (2020). Development of therapeutic antibodies for the treatment of diseases. *Journal of biomedical science*, 27(1), 1–30.

Nikdel, A., Braatz, R.D., and Budman, H.M. (2018). A systematic approach for finding the objective function and active constraints for dynamic flux balance analysis. *Bioprocess and biosystems engineering*, 41, 641–655.

Appendix A. METABOLIC KINETIC CONSTRAINTS

$$\mathrm{f}_{X_v} = K_{01}X_v\left(\frac{[Glc][Gln]}{(K_{02}+[Glc])(K_{03}+[Gln])}\frac{1}{1+\frac{[Amm]}{K_{04}}}\frac{1}{1+\frac{[Glu]}{K_{05}}}\right) - K_{06}X_v^2\left(\frac{[Amm]}{K_{07}+[Amm]}\frac{Glc}{K_{08}+[Glc]}\right)$$

$$f_{Ala} = \left(\left(\frac{K_{10}[Glc]}{K_{11}+[Glc]} + \frac{K_{12}[Gln]}{K_{13}+[Gln]} + \frac{K_{14}[Ser]}{K_{15}+[Ser]}\right)\left(\frac{[Glu]}{K_{16}+[Glu]}\right) - \frac{K_{17}[Ala]}{K_{18}+[Ala]}\right)$$

$$f_{Amm} = \left(\frac{K_{21}[Gln]}{K_{22}+[Gln]} + \frac{K_{23}[Gly]}{K_{24}+[Gly]} + \frac{K_{25}[Asn]}{K_{26}+[Asn]} - \frac{K_{27}[Amm][Glu]}{(K_{28}+[Amm])(K_{29}+[Glu])} - \frac{K_{30}[Amm]}{K_{31}+[Amm]}\right)$$

$$f_{Arg} = \left(-\frac{K_{32}[Arg]^{K_{33}}}{K_{34}+[Arg]^{K_{33}}}\right)$$

$$f_{Asn} = \left(\frac{K_{41}[Asp][Gln]}{(K_{42}+[Asp])(K_{43}+[Gln])} - \frac{K_{44}[Asn]^{K_{45}}}{K_{46}+[Asn]^{K_{45}}}\right)$$

$$f_{Asp} = \left(\frac{K_{51}[Asn]}{K_{52}+[Asn]} - \frac{K_{53}[Asp]}{K_{54}+[Asp]}\right)$$

$$f_{For} = \left(\frac{K_{61}[Trp]}{K_{62}+[Trp]} - \frac{K_{63}[For]}{K_{64}+[For]}\right)$$

$$f_{Glc} = \left(-\frac{K_{71}[Glc]^{K_{72}}}{K_{73}+[Glc]^{K_{72}}}\right)$$

$$f_{Gln} = \left(\frac{K_{81}[Amm][Glu]}{(K_{82}+[Amm])(K_{83}+[Gln])} - \frac{K_{84}[Gln]}{K_{85}+[Gln]}\right)$$

$$f_{Glu} = \left(\frac{K_{91}[Glc]}{K_{92}+[Glc]} + \frac{K_{93}[Asp][Gln]}{(K_{94}+[Asp])(K_{95}+[Gln])} - \frac{K_{96}[Amm][Glu]}{(K_{97}+[Amm])(K_{98}+[Glu])} - \frac{K_{99}[Glu]}{K_{100}+[Glu]}\right)$$

$$f_{Gly} = \left(\frac{K_{101}[Ser][Amm]}{(K_{102}+[Ser])(K_{103}+[Amm])} - \frac{K_{104}[Gly]^{K_{105}}}{K_{106}+[Gly]^{K_{105}}}\right)$$

$$f_{His} = \left(-\frac{K_{111}[His]^{K_{112}}}{K_{113}+[His]^{K_{112}}}\right)$$

$$f_{Ile} = \left(-\frac{K_{121}[Ile]^{K_{122}}}{K_{123}+[Ile]^{K_{122}}}\right)$$

$$f_{Lac} = \left(-\frac{K_{131}[Lac]}{K_{132}+[Lac]} + \frac{K_{133}[Gln]}{K_{134}+[Gln]} + \frac{K_{135}[Glc]}{K_{136}+[Glc]}\left(\frac{1}{K_{137}+[Lac]^{140}}\right)\right)$$

$$f_{Leu} = \left(-\frac{K_{141}[Leu]^{K_{142}}}{K_{143}+[Leu]^{K_{142}}}\right)$$

$$f_{Lys} = \left(-\frac{K_{151}[Lys]^{K_{152}}}{K_{153}+[Lys]^{K_{152}}}\right)$$

$$f_{Met} = \left(-\frac{K_{161}[Met]^{K_{162}}}{K_{163}+[Met]^{K_{162}}}\right)$$

$$f_{Phe} = \left(-\frac{K_{171}[Phe]^{K_{172}}}{K_{173}+[Phe]^{K_{172}}}\right)$$

$$f_{Pro} = \left(-\frac{K_{181}[Pro]^{K_{182}}}{K_{183}+[Pro]^{K_{182}}}\right)$$

$$f_{Ser} = \left(\frac{K_{191}[Gly]}{K_{192}+[Gly]} - \frac{K_{193}[Ser]}{K_{194}+[Ser]}\right)$$

$$f_{Suc} = \left(-\frac{K_{201}[Succ]^{K_{202}}}{K_{203}+[Succ]^{K_{202}}}\right)$$

$$f_{Thr} = \left(-\frac{K_{211}[Thr]^{K_{212}}}{K_{213}+[Thr]^{K_{212}}}\right)$$

$$f_{Trp} = \left(-\frac{K_{221}[Trp]^{K_{222}}}{K_{223}+[Trp]^{K_{222}}}\right)$$

$$f_{Tyr} = \left(-\frac{K_{231}[Tyr]^{K_{232}}}{K_{233}+[Tyr]^{K_{232}}}\right)$$

$$f_{Val} = \left(-\frac{K_{241}[Val]^{K_{242}}}{K_{243}+[Val]^{K_{242}}}\right)$$

X_v is the viable cell density (1e6 cells/mL), and the concentrations of metabolites (mM) are denoted as follows: [Ala]-alanine, [Amm]-ammonia, [Arg]-arginine, [Asn]- asparagine, [Asp]- aspartate, [For]-formate, [Glc]-glucose, [Gln] -glutamine, [Glu]- glutamate, [Gly]-glycine, [His]-histidine, [Ile]-isoleucine, [Lac]-lactate, [Leu]-Leucine, [Lys]-lysine, [Met]-methionine, [Phe]-phenylalanine,[Pro]-proline , [Ser]-serine, [Succ]-succinate, [Thr]-threonine, [Trp]- tryptophan, [Tyr]-tyrosine, and [Val]-valine. The kinetic parameters are given by K_{01} to K_{244}.